Integrating Models of Attention, Numerical Perception, and Quantified Language Generation

One challenge for future interactive natural language generation (NLG) systems is the fact that assumptions of perfect, a priori knowledge do not apply to most real-world situations. Knowledge about the world is instead incrementally acquired through temporally-extended and attentionally-dependent perceptual processes. For example, Barr, Deemter, and Fernández (2013) found that when individuals produce quantified reference expressions (QREs), the form of QRE was dependent on the numerosity of the sets under consideration. When the quantities in each set were large, people tended to produce relational expressions (e.g., “my set is the largest one”). However, when the target set of objects consisted of a small quantity within what is known as the subitizing range, people tended to produce QREs with exact numerical descriptors despite non-exact QREs being sufficient to disambiguate the expression. They concluded that Grice's Maxim of Quantity (Grice, 1975) appears to be violated with respect to over-informativeness in these cases, and that a principle of perceptual effort was a better explanation for their data. Thus, understanding and modeling the psychophysics of human perceptual processes is often necessary to account for human linguistic behavior.

We report on both a novel computational model of numerical perception and description, as well as psychological studies that further clarify the relationship between numerical language and perception. The perception of numerosity consists of multiple processes. Explicit *counting* provides a slow but precise determination of number (Gallistel and Gelman, 1986), while *estimation* provides a rapid but less precise judgment of the quantity of a group of objects (Barth, Kanwisher, and Spelke, 2003). Between these two procedures, a third process, called *subitizing* provides both rapid and precise judgments of numerosity, but only for small quantities, typically around four objects (Kaufman et al., 1945). The processes involved in the perception of numerosity each yield mental representations of numerosity of differing precision. The precision of these mental representations of number in turn constrains and influences the numerical language used to describe a visual scene. Briggs, Bridewell, and Bello (2017) developed a computational model of numerical perception that operates by first ascertaining an approximate, noisy estimate of quantity (e.g. Barth et al., 2003), which is subsequently refined by slower, serial subitizing and counting processes. We extended this model to produce inexact numerical descriptions of quantity (specifically, hedged numbers and intervals) when the resultant numerical representation had some level of uncertainty.

To investigate the effect of numerosity and perceptual time constraints on quantified language, we conducted two experiments. Participants, recruited from Amazon Mechanical Turk, were presented with nine videos. Stimulus presentation duration and the quantity of elements (dots) presented, were manipulated. Three possible presentation durations were used: 200 ms, 1000 ms, and an indefinite amount of time (dots remained on the screen while subjects responded). Three possible quantity ranges were used: 1-4, 5-8, and 9-12. Participants were presented with one video each from each of these duration/quantity conditions in a random order. For each video, participants were asked to complete the following sentence, “In the video above, there are ______ black dot(s).” In the first experiment, we also asked participants to report their confidence in their completed description (1 = very unsure to 5 = very confident). Descriptions were coded by two annotators and classified as being either exact numbers (e.g., “four,” “seven,” “a dozen”) or a type of inexact quantifier. Inexact quantifier classifications include hedged numbers (e.g., “about four”), intervals (e.g. “more than four,” “less than ten,” “between five and eight”), or vague quantifiers (e.g. “a few,” “several,” “many”). High inter-annotator agreement was found (Cohen's kappa = .945).

Like Barr and colleagues (2013), our model predicted that the limits of human perceptual performance would influence language usage. Specifically, we hypothesized that: (H1) as presentation duration decreased, the frequency of non-exact quantifiers would increase; (H2a) in the subitizing range (quantity 1-4) subjects would use exact quantifiers; (H2b) for indefinitely long presentation durations, subjects would use exact quantifiers; and (H3) in difficult duration/quantity pairings, subjects that responded by describing quantity using non-exact quantifiers would indicate *higher* confidence in their response vs. subjects that responded in the same duration/quantity condition with exact numbers.

All hypotheses were supported by the data. For H1, Chi-square tests of independence indicated significant differences in response type frequencies between presentation time categories for quantities 5-8

and 9-12 (p<.001). Only exact descriptions were used when participants had an unlimited amount of time to enumerate (H2b), and the vast majority of descriptions (>90%) in the subitizing range were exact regardless of presentation time condition (H2a). Finally, Wilcoxon-signed rank tests indicated confidence ratings were significantly lower for exact responses than inexact responses for quantity ranges 5-8 (p=.039) and 9-12 (p<.001) at 200 ms presentation time and quantity range 9-12 (p=.007) at 1000 ms presentation time (H3). Interestingly, only 59% of subjects used inexact quantifiers. The other 41% only used exact numbers, guessing in cases of uncertainty. In our second experiment, we investigated whether or not removing the confidence question would affect the distribution of exact/inexact responses. One possibility was that subjects were more inclined to guess exact numbers if they had the option to hedge in a separate confidence question. However, chi-square tests of independence comparing counts of exact vs. inexact quantifiers did not indicate a significant effect of the confidence question (18.3% vs. 20.3% inexact responses, with and without confidence question, respectively). As in the first experiment, all hypotheses were supported.
The results of the human-subjects data were highly correlated with the proportional of exact/inexact quantifiers selected by our computational model (Spearman's rho = 0.938). Replicating the proportion of guessers (exact-only responders) and inexact quantifier users (41/59%) increased model fit even more (Spearman's rho = 0.956).

Future work is planned to adapt this model of temporally-extended numerical perception toward generating QREs. The first step is modeling the data gathered by Barr, Deemter, and Fernández (2013) in a human-subjects elicitation experiments. In this experiment, human subjects generally produced referring expressions that mentioned either exact number, relative quantity, or the presence or absence of objects. We show that that a straightforward application of the Incremental Algorithm (Dale and Reiter, 1995) on these numerical properties, with a full representation of the visual scene, yields no preference ordering that sufficiently captures the distribution of RE types produced by humans. Using the numerical perception model and the QRE domain, we plan to explore how models of REG could be adapted to operate on partially-encoded and noisy representations of scenes.

References

Barr, D., Deemter, K., & Fernández, R. (2013). Generation of quantified referring expressions: evidence from experimental data. In *Proceedings of the 14th European Workshop on Natural Language Generation* (pp. 157–161).

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, *86*(3), 201-221.

Briggs, G., Bridewell, W., and Bello, P. (2017). A Computational Model of the Role of Attention in Subitizing and Enumeration. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*. (pp. 1672–1677).

Dale, R., & Reiter, E. (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive science*, *19*(2), 233-263.

Gelman, R., & Gallistel, C. R. (1986). *The child's understanding of number*. Harvard University Press.

Grice, H. P. (1975). Logic and conversation. (pp. 41–58).

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. The American Journal of Psychology, 62, 498–525.